DOCUMENT RESUME

ED 304 478 TM 012 880

AUTHOR Case, Susan M.
TITLE How Often is "Often"? The Use of Imprecise Terms in Exam Items.
PUB DATE Feb 88
NOTE 13p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Miami Beach, FL, February 24-27, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Definitions; Higher Education; *Item Analysis; *Licensing Examinations (Professions); *Medicine; *Multiple Choice Tests; National Programs; Physicians; Statistical Distributions; Test Construction; *Testing Problems; Testing Programs
IDENTIFIERS Frequency Data

ABSTRACT

This study was designed to gather data on the meaning of imprecise terms from items written by physicians for their students and by test committees for national licensure and certification examinations. A total of 32 members of test committees who write examination items for various medical specialty examinations participated in the study. Each participant was provided with a list of phrases used in multiple-choice items to express some concept related to frequency of occurrence. Participants were asked to indicate what percentage of time was reflected by each phrase and to express this percentage both as a single number (i.e., 75% of the time) and as a band (i.e., 60 to 80% of the time). The list included 15 terms. Responses of two participants were deleted due to excessive aberrance. Results indicate that the phrases used by item writers to express frequency do not have an operational definition that is commonly shared. Both the single values and phrases varied considerably across individuals. Problems created by vague terms are much more severe for true/false type items than for one-best answer items. These findings support the contention that general guidelines on item writing would be beneficial to the medical community. One table and five figures are provided. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED304478

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Susan M. Case

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HOW OFTEN IS "OFTEN"? THE USE OF IMPRECISE TERMS IN EXAM ITEMS

Susan M. Case, Ph.D.
Senior Evaluation Officer
National Board of Medical Examiners

TM012880

How Often is "Often"? the Use of Imprecise Terms in Exam Items

Susan M. Case, Ph.D.
Senior Evaluation Officer
National Board of Medical Examiners

Introduction

Textbooks on item writing indicate that terms such as often, usually, and frequently should be avoided in multiple choice questions. Physicians who write exam items for their students and those who serve on test committees for national licensure and certification examinations are generally unfamiliar with the educational literature and use imprecise terms such as these with great regularity. Item writing workshops that focus on quoting the experts from the educational arena have little impact in changing the behavior, at least in part because physicians believe the terms do have a common definition among practitioners. The language that they use in their items reflects language used in medical discussions (eg, "Obesity is frequently associated with hypertension").

The purpose of this study was to gather data on the meaning of imprecise terms from the item writers themselves. If consensus about the meaning of these terms was found among physicians, then the general guidelines related to item writing would not apply to the specific area of medical education and evaluation. If, on the other hand, little consensus on the meaning of the terms was found, then the data could be used as an illustration to support the contention that the guidelines are applicable to the medical community.

Method

A total of 32 members of test committees who write examination items for various medical specialty examinations participated in the study. Each participant was presented with a list of phrases used in multiple choice questions to express some concept related to frequency of occurence. They were asked to indicate what percentage of time was reflected by each phrase and to express this percentage both as a single number (eg 75% of the time) and as a band (eg, 60 to 80% of the time). Fifteen terms were included on the list. The responses of two individuals were deleted from analysis because they were so aberrant.

Presented at the Eastern Educational Research Association (EERA) Annual Conference, Miami Beach, February 25, 1988. Please direct questions to Susan M. Case, Ph.D., Senior Evaluation Officer, National Board of Medical Examiners, 3930 Chestnut Street, Philadelphia, PA 19104.

TM012880

Results

Table 1 shows the mean value and standard deviation of the single number associated with each term on the list. The means ranged from a high of 84% of the time for the phrase "most of the time" to a low of 5% of the time for the phrase "almost never". There was considerable variation in the value listed for each term; standard deviations ranged from a low of 5.5 to a high of 21.7. The values listed for over half of the phrases spanned 50 percentage points.

Figures 1 and 2 show the distribution of responses for two of the phrases, "most of the time" and "often", selected to serve as illustrations of the data. A total of 40% of the respondents defined "most of the time" as 90% of the time while 30% defined it as 80% of the time. Despite this difference of opinion, this is more concensus than was found in most of the other phrases. Overall, the distribution of values used to define the phrase ranged from a low of 60% of the time to a high of 99% of the time. The mean was 84%.

In contrast, approximately 20% of the respondents defined "often" as 60% of the time; an additional 20% defined it as 70% of the time. More than half of the respondents listed other values, ranging from a low of 20% to a high of 90%, with a mean of 60%.

Figure 3 shows box-plots associated with each term. The mean is the center of the box; the top and bottom of the box are 1 SD from the mean. The lines are extended to the lowest and highest value listed.

The overlap between the terms was considerable. Terms with a broad range of values included "commonly" with a range from 20 to 90; "frequently" with a range from 20 to 95; and "likely to occur" with a range from 25 to 95. The phrase with the most concensus was "approximately half" which was defined as 50% of the time by 87% of the respondents.

There was little concensus regarding the bands for the terms. Figures 4 and 5 show the bands used by respondents to define the terms used for illustration above, "most of the time" and "often". The phrase "most of the time" showed more consensus than most of the other phrases. The band used most often to define the phrase was 70-90% of the time (20% of the respondents). Three other bands were listed by 13% of the respondents: 75-99%, 80-95%, 85-95%.

Only one of the bands for the term "often" was listed by as many as 13% of the respondents: 60-81%. Only two additional bands was listed by more than one person: 70-80% and 50-70%.

Discussion

The results indicate that the phrases used by item writers to express frequency do not have an operational definition that is commonly shared. The mean value plus or minus one standard deviation typically exceeded 25 percentage points.

Not only did the single values vary considerably among individuals, but the bands assigned to the phrases varied considerably as well. Phrases that had precise definitions (narrow bands) for one individual were not necessarily those that had precise definitions for the others. In addition, for some of the phrases (eg "often"), phrases were defined rather precisely, but there was little overlap among individuals in the values included in the band.

It should be noted that imprecise terms are used in our everyday speech and in our writing. Many of the phrases used in the study are also used in the text of this paper, without (it is hoped) creating any confusion. However, imprecise terms may cause confusion when they are used in the text of examination items.

These results have different implications for the various item formats. Problems created by vague terms are much more severe for true-false type items (K-, C- and X-type items) than for one-best answer (A- and B-type) items.

For example, imprecise terms cause major problems in multiple true-false items such as that listed below:

True statements about cystic fibrosis (CF) include:

1. CF is associated with short stature.
2. Children with CF usually die in their teens.
3. Males with CF are often sterile.
4. CF primarily affects the lungs.

Modifying the item by specifying exact numbers doesn't correct the problem. For example, the statement "the incidence in the US is 1:2000" can not be judged as true or false. Making it more vague by stating "the incidence in the US is approximately 1:2000" does not help since the band is not specified. In true/false items, the appropriate treatment of numerical items is either to generate a comparison (eg, the incidence of CF is greater than that of juvenile diabetes) or to specify a range (eg, the incidence of CF is greater than 1:1500).

The problem noted above with multiple true-false items is not as acute with well-constructed "one-best answer" items (ie, those that pose a clear question). For example, the following has a vague term in the stem, yet because the task is to select the one-best answer, the item is relatively unambiguous.

Children born with CF are most likely to die

A. before the age of 1
B. between the ages of 1 and 5
C. between the ages of 5 and 10
D. when they are teenagers
E. when they are in their 20s

Problems do arise with one-best answer items like the following:

Children with CF have problems with their digestive systems

A. frequently
B. usually
C. often
D. most of the time

The only way to make such an item worse is to use a fifth option "none of the above".

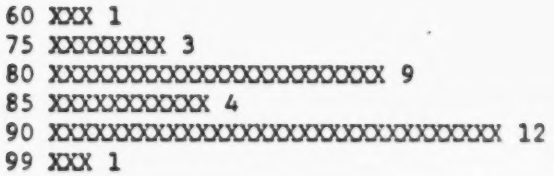
TABLE 1

Summary Data
for Frequency Terms

Terms	Mean	SD
Most of the time	84.1	7.5
Most likely to occur	78.3	12.5
Primarily	77.9	13.1
Most often	72.7	15.0
Usually	71.9	10.4
Likely to occur	65.8	16.1
Probably	64.9	11.5
Commonly	62.6	16.2
Frequently	62.4	19.2
Often associated with	61.0	15.7
Often	60.5	18.4
Associated with	59.5	21.7
Approximately half	51.9	8.6
Rarely	8.2	5.5
Almost never	5.1	5.7

Figure 1

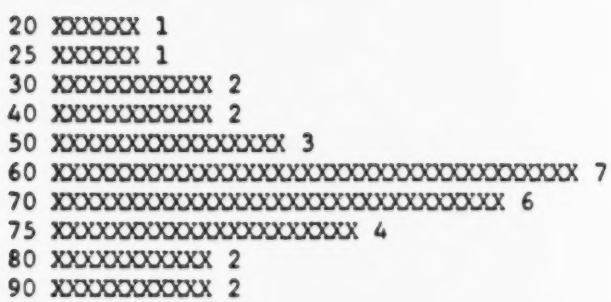
Distribution of the responses defining the phrase "Most of the time"



Mean	84.133	Std Err	1.366	Median	85.000
Mode	90.000	Std Dev	7.482	Variance	55.982
Kurtosis	2.426	S E Kurt	.833	Skewness	-.994
S E Skew	.427	Range	39.000	Minimum	60.000
Maximum	99.000	Sum	2524.000		

Figure 2

Distribution of the responses defining the phrase "Often"



Mean	60.500	Std Err	3.359	Median	60.000
Mode	60.000	Std Dev	18.399	Variance	338.534
Kurtosis	-.218	S E Kurt	.833	Skewness	-.623
S E Skew	.427	Range	70.000	Minimum	20.000
Maximum	90.000	Sum	1815.000		

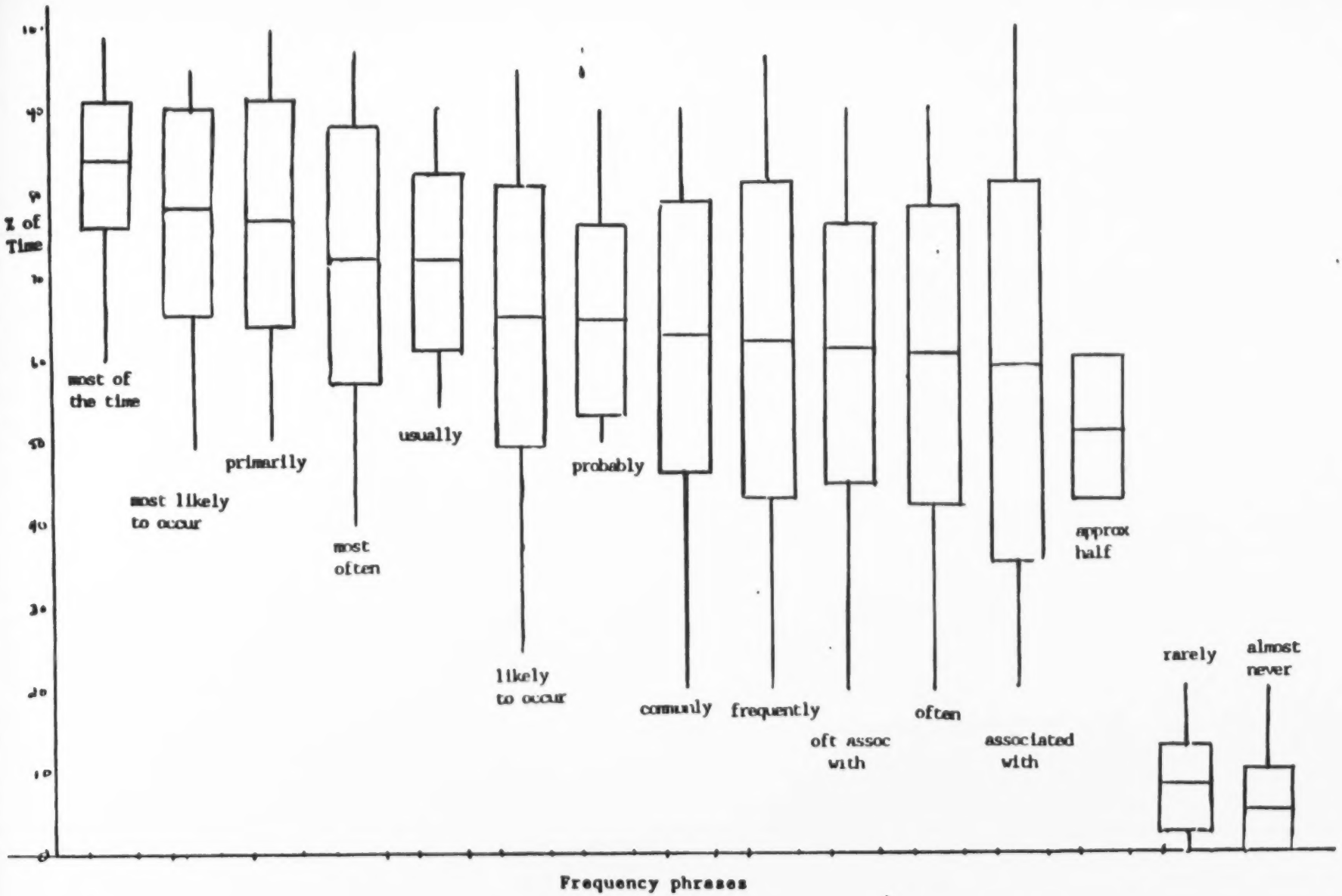
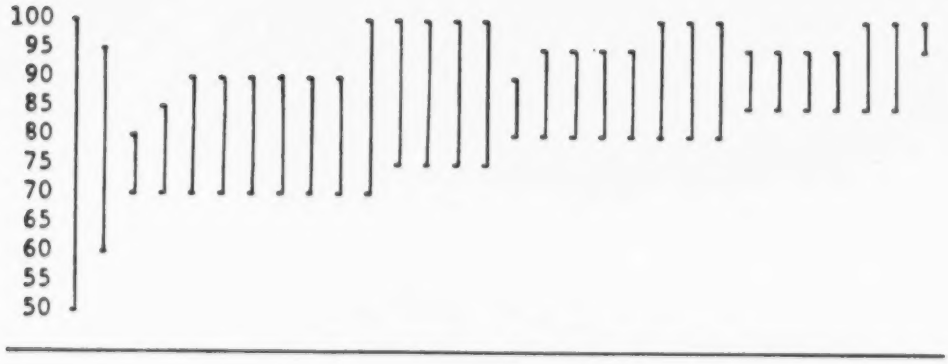


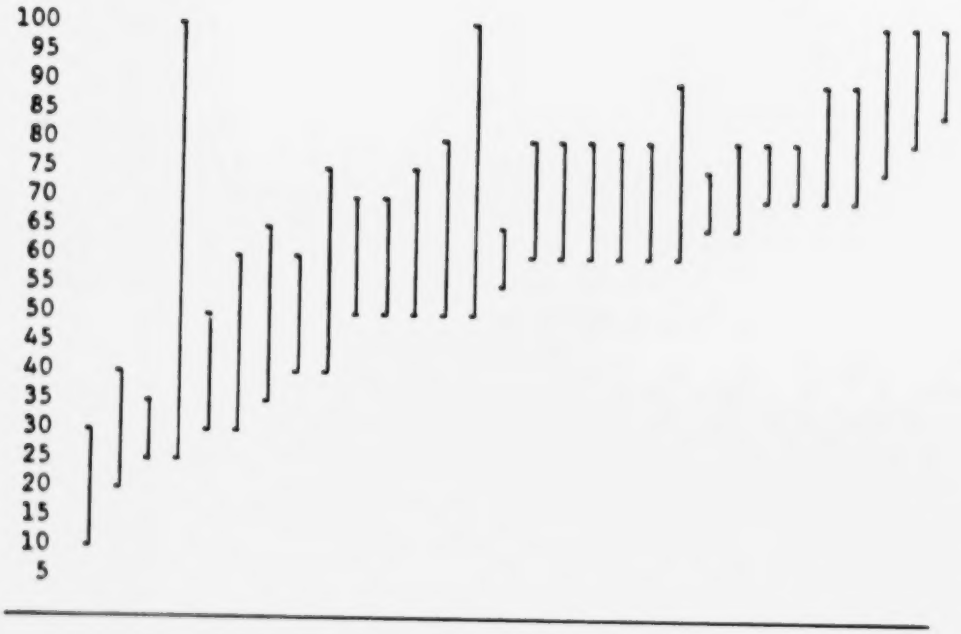
Figure 3. Box plots showing distribution of responses for frequency terms

Figure 4
Bands Listed for
"Most of the Time"



Band	n
51-99	1
60-95	1
70-80	1
70-85	1
70-90	6
70-99	1
75-99	4
80-90	1
80-95	4
80-99	3
85-95	4
85-99	2
95-99	1

Figure 5
Bands Listed for "Often"



Band	n	Band	n
10-30	1	50-99	1
20-40	1	55-65	1
25-35	1	60-80	4
25-99	1	60-81	1
30-50	1	60-90	1
30-60	1	65-75	1
35-65	1	65-80	1
40-60	1	70-80	2
40-75	1	70-89	1
50-70	2	70-90	1
50-75	1	75-99	1
50-80	1	80-00	1
		85-99	1